Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Acoustics xxx (2005) xxx–xxx
applied acoustics
www.elsevier.com/locate/apacoust

Prediction of speech intelligibility for public address systems in traffic tunnels

Sander J. van Wijngaarden *, Jan A. Verhave

TNO Human Factors, PO Box 23, 3769 ZG Soesterberg, The Netherlands

Received 1 December 2004; received in revised form 1 July 2005; accepted 8 July 2005

Abstract

Traffic tunnels are generally hostile acoustic environments, both in terms of reverberation and ambient noise levels. Public address (PA) systems used to convey spoken warnings must meet stringent design requirements in order to produce sufficiently intelligible speech. To be able to predict PA system performance at tunnel design time, two different speech transmission index (STI) calculation procedures were implemented. The first procedure predicts the STI based on ray-tracing simulations. Comparison with measured STI data showed that this simulation approach yields accurate intelligibility estimates. However, the procedure is time-consuming and too complex to be used by non-specialists. For this reason, a second (simpler and more efficient) procedure was developed, based on fixed non-linear regression, statistically deriving prediction functions from measured data and ray-tracing results. This procedure was compared to the approach based on ray tracing, and found to yield STI predictions closely matching those of the ray-tracing model.

Keywords: Tunnels; Intelligibility; Prediction; STI

---

* Corresponding author.
E-mail address: sander.vanwijngaarden@tno.nl (S.J. van Wijngaarden).

doi:10.1016/j.apacoust.2005.07.001

1. Introduction

During a traffic tunnel design process, many design criteria tend to prevail over acoustic considerations. The geometrical design is inherently disadvantageous; long, pipe-like enclosures such as tunnels are likely to show relatively long reverberation times. Acoustic absorption of surfaces is usually also less than optimal. These surfaces are designed for their ability to withstand prolonged heat exposure, as may occur in case of a severe fire. Consequently, walls, road surface and ceiling are acoustically relatively hard. Even when a compromise between heat resistance and acoustic absorption appears feasible (application of acoustic plasters, or using fireproof sound absorbent materials to cover the walls), cost considerations and problems associated with cleaning are likely to prevent 'soft' materials from being used [1].

In addition, noise sources may be expected to produce high sound levels in the tunnel, partly due to the reinforcing effect of reverberation. The main sources of noise are moving traffic and high-power ceiling-mounted ventilators, used to expel exhaust fumes or smoke.

1.1. Public address systems in tunnels in the Netherlands

Hostile tunnel acoustics are especially problematic to the performance of public address (PA) systems. A widely used standard solution, horn-type loudspeakers attached to the ceiling in arrays, may or may not offer sufficient intelligibility; this not only depends on the tunnel characteristics, but also on the PA system design. Factors such as loudspeaker directivity, linearity, maximum sound power, frequency response and propagation delay correction are very important to the performance of the system. A schematic representation of a typical PA system as used in traffic tunnels in the Netherlands is shown in Fig. 1.

Spoken messages are produced in a traffic control centre, usually at a distance from the tunnel. This implies that the operator (either using his own voice or prerecorded messages) does not receive auditory feedback; he or she has no means to determine whether the messages are intelligible to the public inside the tunnel.

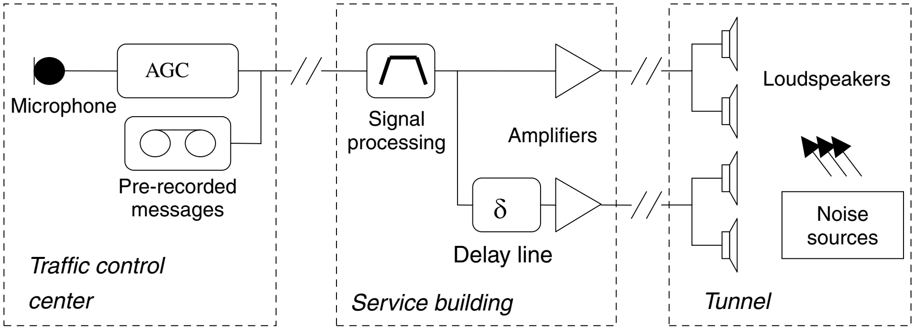

Fig. 1. Schematic representation of a PA system in a traffic tunnel. AGC stands for automatic gain control.

Since traffic tunnels are generally too long to address all at once, they need to be subdivided into sections (typically 120 m in length). The operator can remotely view the tunnel, section by section, through a closed-circuit television (CCTV) system. The PA system only produces sound in the section currently activated by the operator (i.e., the section currently shown on the CCTV screen). If the operator wants to address the entire tunnel, this has to be done one section at a time.

In each section, typically two or three arrays of loudspeakers are present. Delay lines are used to compensate travel time differences, preventing contributions from consecutive arrays to be perceived as discrete echoes. Additional signal processing may be used to enhance the overall performance of the system.

1.2. Measurements of speech intelligibility in traffic tunnels

In the Netherlands, PA systems in traffic tunnels are required to meet minimum requirements with regard to speech intelligibility. Applicable regulations require that speech intelligibility is quantified by means of the speech transmission index (STI) [2,3]. This method (described in more detail in Section 2 of this paper) uses artificial test signals to measure the degree to which intensity modulations, as present in natural speech, are preserved while sound propagates towards a listener. Since preservation of these modulations implies preservation of speech transmission quality [4], the STI is a reliable predictor of speech intelligibility for many applications.

Whenever a new tunnel is built in the Netherlands, or the PA system in an existing tunnel is replaced, a standardised STI measurement protocol is applied to verify compliance with the intelligibility requirements. Such a compliance test can be carried out using STI-measuring devices and software available from a number of different vendors. Out of several IEC-standardised test signal types, the STI-PA test signal is the most suitable for the application described in this paper [5]. A specific test protocol for measuring STI measurements in tunnels is used.

As part of the test protocol, representative and worst-case sections of the tunnel are selected for carrying out measurements. For each section, STI measurements are carried out at a set of pre-defined positions. There positions are normally 15 m apart (measured along the length of the tunnel), on each lane, at a height of 1.50 m (see Fig. 2).

STI results normally turn out worse than average in the proximity of noise sources (such as ventilators) and at positions further away from the loudspeakers. Positions further away from a loudspeaker array, but close to the next (delayed) loudspeaker array, are normally also among the ones producing the lowest STI values. Specific acoustic features in the tunnel, such as reflection signs and lights, also locally decrease the STI.

Unfortunately, the compliance test procedure described above can only be carried out *after* the tunnel is built and the PA system has been taken into operation. For the optimisation of PA system performance, it would be much better to evaluate the system throughout the various design stages, when changes can still be made without resulting in significant additional costs. Since it is impossible to *measure* performance at this stage (since the PA system, or even the tunnel, does not exist yet), *predictive calculations* of the STI are needed. For this reason, a prediction model for speech intelligibility based on

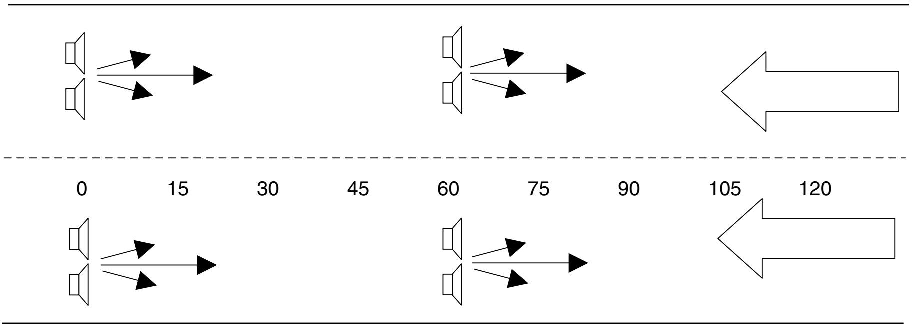

Fig. 2. Example of an arrangement of loudspeaker arrays in a two-lane traffic tunnel, with a section length of 120 m. The traffic in both lanes is travelling in the same direction, indicated by the large arrows. Clusters of loudspeakers are mounted on racks attached to the ceiling, above the middle of each lane, and aimed toward the moving traffic. At the position marked "0" the signal is presented without delay; the sound reproduced by the clusters 60 m further away is delayed.

acoustic ray tracing has been implemented. Ray tracing has been shown capable of yielding accurate STI estimates in industrial halls [6] and in rooms in general [7]. Our implementation targeted at traffic tunnels was validated by comparison with measurement results in existing tunnels. The applied prediction procedure and an evaluation of the accuracy of this approach are described in this paper.

Experience with the ray-tracing approach showed that its accuracy is, for some purposes, offset by its complexity. An important drawback is the necessity to construct a specific geometrical computer model of the tunnel interior, for each individual tunnel (or even tunnel section) that is to be evaluated. Also, we found that designers without specific expertise in intelligibility prediction found the procedure too complicated for practical use. More specifically, obtaining (or estimating) the input parameters for the ray-tracing model can be a problem. Inaccurately estimated input parameters will result in faulty predictions.

Theoretically, the STI can also be calculated on the basis of a much simpler empirical statistical procedure. By using actual STI measurement data and results from reliable ray-tracing predictions, and deriving prediction functions from this data using fixed non-linear regression, relatively straightforward computational procedure can be derived. Similar procedures have previously been successfully applied to obtain estimates of noise levels, reverberation times and the STI in industrial workrooms [8] and in classrooms [9,10]. A regression-based procedure for STI prediction in tunnels is also presented in this paper, as well as its validation.

2. Calculation of the speech transmission index

2.1. The speech transmission index method

The speech transmission index (STI) gives an objective prediction of speech intelligibility by means of a single 0–1 index [2,3]. The STI has been validated through comparison between STI scores and subjective speech intelligibility scores, using various intelligibility assessment methods in many different conditions [11,12].

In an STI score all relevant types of speech degrading influences encountered in tunnel acoustics (such as noise, echoes and reverberation) are incorporated. Signal degradation due to limitations of the performance of a PA system (such as non-linear distortion components) are also taken into account.

As stated in Section 1, the STI is based on the principle that preservation of intensity modulations implies preservation of speech. The degree to which these intensity modulations are preserved is expressed by the modulation transfer function (MTF) intelligibility [13,14]. The STI can always be derived from the MTF through a standardised [2] and straightforward procedure (see Section 2.2). Mathematically, the challenge to accurately measure or predict the STI essentially comes down to finding an accurate estimate of the MTF.

In order to *measure* the MTF, a number of alternative approaches can be adopted, which all have in common that the intensity modulations at the output side of the channel (a listener position in the tunnel) are compared to the modulations present in the test signal, as applied to the input of the channel. In order to be able to *predict* (calculate) the MTF, a sufficiently accurate mathematical model of the transmission channel is needed. In Sections 2.3 and 2.4, two alternative approaches for defining such a model are described.

In Table 1, the correspondence between STI values and a qualification of speech intelligibility is given, together with a well-known subjective intelligibility measure (equally balanced open-set CVC word score).

Predictive calculations of the STI are only useful if they can be made with sufficient accuracy: given a set of representative conditions, the difference between measured STI and predicted STI scores must be small relative to the STI intervals that determine the five qualification categories in Table 1. This criterion shall be used to evaluate the STI prediction procedures proposed in this paper.

2.2. Basic STI calculation procedures

The basic STI algorithms, as standardised in IEC 60268-16 3rd edition [2], pivot around the fact that preservation of intensity modulations can be quantified by means of a modulation index m. Modulation index m can be measured using a test

Table 1
Qualification and relation between STI and CVC word score

Qualification	STI	CVC word score (% correct)
Excellent	>0.75	>96
Good	0.60–0.75	86–96
Fair	0.45–0.60	65–86
Poor	0.30–0.45	32–65
Bad	<0.30	<32

The CVC-word score was obtained in an open-response paradigm, using equally balanced CVC-word lists [7].

signal with (mean) intensity $\bar{I}_\mathrm{s}$, which has been designed to have a modulation depth of 100%. For the moment, we will assume that some arbitrary modulation frequency is used. If a (non-modulating) noise signal with intensity $\bar{I}_\mathrm{n}$ is introduced on the channel, then this will reduce the modulation depth of the overall signal at the output of the channel. While the signal was originally modulated across 100% of the overall intensity range, at the output of the channel this modulation will now no longer cover the entire intensity range of the signal. The reduction of the modulation range is expressed by modulation index m, and given by

$$m = \frac{\bar{I}_\mathrm{s}}{\bar{I}_\mathrm{s} + \bar{I}_\mathrm{n}}. \tag{1}$$

If the signal degradation due to the channel consists of additive noise only, a measured modulation index m has a one-to-one relation with the signal-to-noise ratio. If the modulation depth is also reduced through causes other than additive noise (such as reverberation), then a modulation index m can still be measured, but the value of m now also incorporates degradations other than noise. Using Eq. (1), and by logarithmic conversion of the signal and noise intensities onto the decibel scale, any modulation index m can still be translated into an *effective* signal-to-noise ratio $\mathrm{SNR}_\mathrm{eff}$

$$\mathrm{SNR}_\mathrm{eff} = 10 \log \frac{m}{1-m}\ \mathrm{dB}. \tag{2}$$

Through Eq. (2) the STI model relates the intelligibility effects of all sorts of speech signal degradations to the effect of adding noise at an *equivalent* speech-to-noise ratio.

The nature of many speech transmission channels is such, that m may very well depend on modulation frequency F. Also, m will often vary as a function of signal frequency itself. If the signal is analysed in octave bands (indicated by an octave band index n), then by definition the modulation transfer function (MTF) is $m(n, F)$ [15].

The MTF can be measured or calculated in various ways; the form in which the MTF is presented differs. In standardised STI calculations, the MTF appears as a discrete 14×7 matrix (modulation index for 14 modulation frequencies in 7 frequency octave bands).

$$m(n, F(i)) = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,n} & \cdots & m_{1,7} \\ m_{2,1} & \ddots & & & & \vdots \\ \vdots & & \ddots & & & m_{i,7} \\ m_{i,1} & & & m_{i,n} & & \vdots \\ \vdots & & & & \ddots & m_{13,7} \\ m_{14,1} & \cdots & m_{14,n} & \cdots & m_{14,6} & m_{14,7} \end{bmatrix}. \tag{3}$$

The modulation frequencies $F(i)$ included in this matrix represent the 14 1/3-octave centre frequencies from 0.63 to 12.5 Hz, while the 7 octave bands n represent octave bands with centre frequencies ranging from 125 to 8000 Hz.

The next step in the STI calculation procedure is to convert the MTF (Eq. (3)) into effective SNRs using Eq. (2). The STI model assumes that speech intelligibility is linearly related to SNR, across a 30-dB range centred at SNR = 0. Each effective SNR is converted into a so-called transmission index (TI) by mapping this SNR on a 0–1 range according to

$$\mathrm{TI}(n,i) = \frac{\mathrm{SNR}_{\mathrm{eff}}(m(n,F(i))) + 15}{30}. \tag{4}$$

The TI-matrix defined by Eq. (4) quantifies how each modulation frequency $F(i)$ in each octave band n contributes to the overall speech intelligibility. The STI model assumes all modulation frequencies in the 0.63–12.5 Hz range to be equally important. This means that TI-matrix (4) can be simplified to a so-called modulation transfer index (MTI) vector according to

$$\mathrm{MTI}(n) = \frac{1}{14}\sum_{i=1}^{14}\mathrm{TI}(n,i). \tag{5}$$

Eq. (5) results in 7 TI values, one for each octave band. These are combined to produce the STI. Since not all octave bands contribute equally to the overall intelligibility, octave weighting factors α_n are used. Moreover, to incorporate the effect into the calculations that neighbouring frequency bands are mutually dependent [16], so-called redundancy correction factors β_n are introduced

$$\mathrm{STI} = \sum_{n=1}^{7}\alpha_n\mathrm{MTI}_n - \sum_{n=1}^{6}\beta_n\sqrt{\mathrm{MTI}_n \cdot \mathrm{MTI}_{n+1}}. \tag{6}$$

Values of α_n and β_n are somewhat different for male and female speech, and are defined by the IEC standard [2].

In practice, the full calculation of the IEC-standardised STI is slightly more complicated than presented in this section. Although speech intelligibility is primarily determined by (effective) speech-to-noise ratios, the absolute sound levels may become important at low levels (reception threshold) and very high levels (auditory masking). These STI algorithm aspects related to absolute sound level of the speech signal are beyond the scope of this paper. However, it should be noted that IEC-compliant STI-measuring devices [2] *do* incorporate these level-dependencies, and may lead to different (generally lower) STI results than obtained when literally following the approach outlined in this paper. If needed to obtain STI predictions that are more representative of IEC-compliant STI measurements, then the level-dependency of the STI procedure can be added to the procedure described here a relatively straightforward fashion by literally following IEC 60258-16 3rd edition [2].

2.3. Calculation of the MTF from ray-tracing results (impulse responses)

Acoustic simulation procedures, such as acoustic ray tracing, can be used to predict impulse responses between any source and receiver in a simulated room. These impulse responses can then be used to calculate an MTF for this room [17], for the given source and receiver positions. Any simulation requires that at least the following input parameters must be known:

(1) Positions of all sources.
(2) Specification of sources (directivity patterns, sound power).
(3) Geometrical configuration of the tunnel.
(4) Sound absorption coefficients (as a function of octave band) for each surface.
(5) Scatter coefficients for each surface.
(6) (Assumed) positions of listeners in the tunnel, simulation grid.

If effects of air absorption are expected to play a significant role, then temperature and humidity of the air must also be specified.

If a communication channel does not introduce noise or non-linear signal distortions of any sort, then its MTF can be fully derived from its (squared) impulse response [15], according to the procedure described further on in this section. Hence, any standard acoustic model that predicts a room's impulse response can be used for speech intelligibility predictions in that room. In fact, if this impulse response corresponds to a purely exponentially decaying reverberation curve, without contributions from distinct echoes, then the following simplified (and commonly used) equation can be applied for calculating the MTF [15]:

$$m(n,F) = \left[1 + \left(2\pi F \frac{T(n)}{13.8}\right)^2\right]^{-\frac{1}{2}}. \tag{7}$$

Here $m(n,F)$ represents the MTF for a room with octave-band specific T60 reverberation times $T(n)$.

Eq. (7) is commonly, but incorrectly, used by acoustic simulation software for calculation of the overall effects of room acoustics on the STI, ignoring the effects of echoes and reverberation phenomena other than pure exponential decay. Hence, even in the absence of noise and non-linear distortion, Eq. (7) may be inaccurate.

A more generally applicable MTF, although still not including effects of noise and non-linear distortion, can be derived from impulse responses predicted by acoustic simulation procedures [17]. For the experiments described in this paper we used the acoustic ray-tracing implementation of the software package *Odeon 3.1* (*combined edition*), developed by the Technical University of Denmark [18]. In order to balance accuracy with computational complexity, early reflections were calculated using an image source model, whereas late reflections were treated as originating from independent secondary sources. For all our calculations, we set the transition order to 3. All reflections (stemming, according to reflection order, from different calculation models) were combined in joint energy decay curves.

These energy decay curves are, in fact, backward-integrated squared impulse responses giving energy $E(n,t)$ as a function of time t and octave frequency band n. The squared impulse response $r(n,t)$, with n the octave band index and t time, may be derived from the energy curve $E(n,t)$ by using the equation

$$10\cdot{}^{10}\log r(n,t) = \frac{\partial E(n,t)}{\partial t}. \tag{8}$$

Now, according to Houtgast et al. [15], the MTF for modulation frequency F is defined by

$$m(n,F) = \frac{|\int_0^\infty \mathrm{e}^{2\pi jFt} r(n,t)\,\mathrm{d}t|}{\int_0^\infty r(n,t)\,\mathrm{d}t} \tag{9}$$

which is in fact the modulus of the (normalised) Fourier transform of $r(n,t)$. If the energy decay curve would be available as an analytical function, then Eq. (9) could directly be used to calculate an MTF matrix in the form of Eq. (3). As it is, energy decay curves are exported from *Odeon* as discrete tables, with the energy observations in time separated by a known time window (Δt, typically 5–20 ms).[1] In fact, the energy decay curve is stored as a table of $n \times (\frac{t_{\max}}{\Delta t})$ elements, $t_{\max}$ indicating the time span covered by the decay curve. An MTF matrix of the form of Eq. (3) can be calculated numerically, using discrete implementations of Eqs. (8) and (9). This MTF (which we will label $\mathrm{MTF}_{\mathrm{ra}}$) does not (yet) incorporate the effects of noise and non-linear distortions, since it is derived from a room impulse response only (hence the subscript “ra”, which stands for room acoustics). These factors are introduced later on in the calculation procedure. This means that $\mathrm{MTF}_{\mathrm{ra}}$ is different from any MTF obtained through measurements with test signals, which *do* also incorporate effects of noise and non-linear distortions.

Non-linear effects can separately be introduced in STI calculations in the form of an MTF as in Eq. (3), in this case representing *only* modulation transfer effects of non-linear distortions (instead of the effects of rooms acoustics). For public address systems in tunnels the only non-linear behaviour of practical importance is expected from the loudspeakers, which normally operate at high sound levels, close to their operating limits. Since a reliable and sufficiently general prediction model for the non-linear behaviour of loudspeakers under such conditions is lacking, the only way to obtain a quantification of the resulting modulation reduction is through a single STI measurement.

Most STI-measuring devices are capable of outputting full MTF measurement results. By using such an MTF estimate from a single loudspeaker measurement, making sure that no other sources of modulation reduction are affecting the measurement (silent, non-reverberant conditions), a separate MTF ($\mathrm{MTF}_{\mathrm{nl}}$) for non-linear distortions only is obtained.[2] This can be done under laboratory conditions, or even

---

[1] As it is derived from a discrete energy curve, the squared impulse is inherently low-pass filtered. The time window Δt must be sufficiently small (typically < 20 ms), to obtain all modulation frequencies up to 12.5 Hz in the calculated MTF.

[2] Since non-linear distortion is usually independent of modulation frequency, $\mathrm{MTF}_{\mathrm{nl}}$ may normally be represented as a vector of 7 octave bands instead of a 14×7 matrix. However, for reasons of uniformity it will be denoted here formally as $m_{\mathrm{nl}}(n,F)$.

in situ (loudspeaker array already present in a tunnel environment) by placing a measurement microphone close to the loudspeaker array, well within the reverberation radius. It is important that the loudspeakers are operated at representatively high soundlevels, or else the non-linear distortion behaviour may be underestimated.

We have now quantified two separate contributions to the overall speech degradation, each in a separate MTF. A third separate MTF could be calculated to quantify the effect of noise (by using the inverse function of Eq. (2)). The issue is now how to combine these separate MTFs into a single MTF, from which to calculate the STI.

The solution is to expand the "noise" term in Eq. (1) into three separate terms, representing the separate influence of room acoustics, non-linear distortions and additive noise ('real" noise)

$$m = \frac{\bar{I}_{\mathrm{s}}}{\bar{I}_{\mathrm{s}} + \bar{I}_{\mathrm{ra}} + \bar{I}_{\mathrm{nl}} + \bar{I}_{\mathrm{n}}}. \tag{10}$$

Since the intensity $\bar{I}_{\mathrm{s}}$ of the original signal (undistorted speech at the input of the channel) is the same in all cases, the overall modulation index m can be calculated from the (effective) signal-to-noise ratios for the three degradation sources. An MTF $m(n,F)$ for degradation type x can be converted into the intensity domain through the equation

$$\frac{\bar{I}_{\mathrm{s}}(n,F)}{\bar{I}_{x}(n,F)} = \frac{m_x(n,F)}{1 - m_x(n,F)} \tag{11}$$

which is, apart from the decibel conversion, the modulation and frequency dependent inversion of Eq. (2). By combining Eqs. (10) and (11), an overall MTF is calculated

$$m(n,F) = \frac{\bar{I}_{\mathrm{s}}}{\bar{I}_{\mathrm{s}}\left(1 + \frac{1-m_{\mathrm{nl}}(n,F)}{m_{\mathrm{nl}}(n,F)} + \frac{1-m_{\mathrm{ra}}(n,F)}{m_{\mathrm{ra}}(n,F)}\right) + \bar{I}_{\mathrm{n}}(n)}. \tag{12}$$

The intensity of additive noise $\bar{I}_{\mathrm{n}}$ is normally known a priori (or otherwise straightforwardly derived from $\bar{I}_{\mathrm{s}}$ and the speech-to-noise ratio), so this does not have to be derived from a separate MTF calculation.

From Eq. (12), the STI can be calculated according to the standard procedure as described in Section 2.2.

2.4. Simplified MTF calculations based on a regression technique

The procedure described in the previous section is applicable to more than just traffic tunnels; it can be used to obtain MTF estimates in any type of room or enclosure. This comes at a cost: accurately defining the geometrical make-up of a room, with or without computer aided design tools, is always time-consuming. An additional disadvantage is that ray-tracing procedures are computationally expensive, especially if a high degree of detail or accuracy is required.

If the geometrical configuration of the simulated enclosure is always more or less the same, as is the case for traffic tunnels, then an easier and quicker way to calculate the MTF$_{\mathrm{ra}}$ (contribution to the overall MTF related to room acoustics) can be thought of. This procedure is based on defining the acoustic characteristics at a certain cross-section of the tunnel, under the assumption that this cross-section is representative for the entire tunnel. The shortest possible list of variables presumed to be influencing $\mathrm{MTF}_{\mathrm{ra}}$ at this cross-section is compiled (Table 2).

The influence of the acoustic environment can be represented by only two variables: $A(n)$, which is indicative of the volume of the tunnel (but related to a specific cross-section), and $O(n)$ which indicates the amount of acoustic absorption. The only variable related to the PA system that is explicitly included is the directivity of the loudspeaker arrays $Q(n)$. Loudspeaker frequency response, which is also an important determining factor for intelligibility, is addressed implicitly through the inclusion of octave band number n in the regression equations. The listener position is defined using three variables, expressing the distance to the closest loudspeaker array producing useful sound (D), the closest loudspeaker array that is likely to degrade intelligibility due to backward radiation of delayed speech (N) and the absolute distance to the beginning of the section (P). Together, these three parameters can cover a wide range of different loudspeaker section configurations.

Frequency is included explicitly in Table 2, both linearly and in octave bands, to cover the possibility that some acoustic phenomena depend linearly on frequency and others logarithmically.

Next, we need to define *how* we assume that $\mathrm{MTF}_{\mathrm{ra}}$ depends on the variables of Table 2. Rather than trying to predict the relation between these variables and the MTF directly, we do this for the effective speech-to-noise ratio (in dB) related to room acoustics, $S_{\mathrm{ra}}(n)$.

$$\begin{aligned} S_{\mathrm{ra}}(n) = {} & a_{(0)} + a_{(1)}A(n) + a_{(2)}O(n) + a_{(3)}Q(n) + a_{(4)}n + a_{(5)}\log\left(\frac{D}{15}+1\right) \\ & + a_{(6)}D + a_{(7)}P + a_{(8)}\left(\frac{N}{Q(n)}\right)^{a_{(9)}} + a_{(10)}f(n) + a_{(11)}N. \end{aligned} \tag{13}$$

Table 2
Parameters presumed to be related to the intelligibility of PA systems in tunnels, along with their symbols and units as used in the equations

Symbol	Unit	Range	Parameter
n	–	1–7	Octave band index, ranging from one (125 Hz) to seven (8 kHz)
$f(n)$		125–8000	Centre frequency (125 Hz – 8 kHz) of each octave band
$A(n)$	m^2	50–150	Area of the tunnel cross-section at a given listener position
$O(n)$	m	0.5–15	Equivalent "open window" area due to acoustic absorption, for each meter of tunnel length at the current listener position
$Q(n)$	–	1–18	Q-factor of each loudspeaker (assumption: only loudspeakers of the same type are used)
D	m	−15 to 50	Distance to the nearest loudspeaker array *pointing towards the listener*
P	m	−15 to 200	Distance to the beginning of a loudspeaker section (distance to the first loudspeaker array in the section)
N	m	−100 to 150	Distance to the nearest loudspeaker array *pointing away from the listener*

The range of values that may realistically occur in traffic tunnels is also given for each variable.

Eq. (13) shows that we assume that 11 parameters are needed to predict $S_{\mathrm{ra}}(n)$ and (through dB-to-linear conversion and Eq. (11)) $\mathrm{MTF}_{\mathrm{ra}}$. $S_{\mathrm{ra}}(n)$ is expected, based on relationships between the regression variables, to depend approximately linearly on all variables. However, for distance D there is also component for which a logarithmic relation is predicted, while for the contribution of distance N (normalised for directivity $Q(n)$) a power law relation is expected. These fixed non-linear terms were included to enable the formula to mimic the approximate position-dependence observed in real STI measurements. Notice that by the form of Eq. (13) octave bands are presumed to be mutually independent.

Using fixed non-linear multiple regression [19], values of the parameters $a(k)$ can be derived statistically from measurements of $S_{\mathrm{ra}}(n)$.

Assuming a similar, slightly simplified, relation for the absolute speech level $L(n)$,

$$\begin{aligned} L(n) = {} & b_{(0)} + b_{(1)}A(n) + b_{(2)}O(n) + b_{(3)}Q(n) + b_{(4)}n + b_{(5)}\log\left(\frac{D}{15}+1\right) \\ & + b_{(6)}D + b_{(7)}P + b_{(8)}N^{b_{(9)}} + b_{(10)}f(n) \end{aligned} \tag{14}$$

this level may predicted using the same procedure, once parameters $b(k)$ have also been derived through fixed non-linear regression.

Once $\mathrm{MTF}_{\mathrm{ra}}$ is calculated through Eq. (13), then the rest of the calculation is the same as for the ray-tracing-based approach presented in Section 2.3.

3. Validation of the STI calculation procedures

3.1. Simulated tunnel environments

Geometrical models of the interiors of three existing tunnels (referred to from now on as A, B and C) were constructed. All three tunnels were built from pre-fab concrete segments, and had geometrically similar cross-sections (see Fig. 3). The three modelled tunnels differed in height, length, width, curvature, acoustic absorption characteristics, and configuration of the loudspeaker clusters. For one tunnel (tunnel A), two different loudspeaker section configurations were modelled (A1 and A2). This makes the total number of different configurations four. These four configurations had all been physically realised in existing tunnels, for which elaborate STI and MTF measurements were made.

The tunnel geometry was translated from blueprints to the co-ordinate system required by *Odeon*. All tunnels were curved in both the horizontal and vertical directions. This was modelled by splitting the tunnel up in segments, which were placed at suitable angles to approximate the curvature.

In all four simulated situations, loudspeakers of the horn-type were mounted in brackets attached to the ceiling. The loudspeakers were horizontally pointed straight ahead, facing the oncoming traffic. In the vertical direction, the loudspeakers were slightly tilted towards the road surface, placed at an angle of 3° relative to the ceiling.

The input parameters required for the ray-tracing procedures were taken from blueprints, loudspeaker specifications and material databases. These parameters

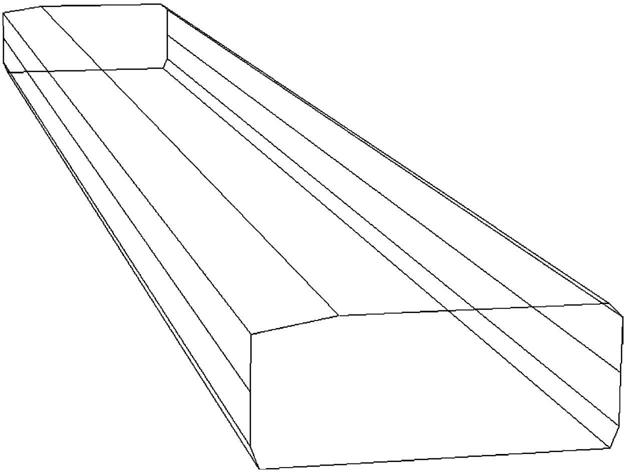

Fig. 3. Example of the 3D view of a tunnel segment. The ceiling consists of three separate surfaces, since the cross-section of the tunnel is not perfectly rectangular. Because upper and lower parts of the walls usually have different surfaces materials, the wall surfaces are also built up from multiple surfaces. The road surface has a slight angle relative to the central part of the ceiling.

were not changes after inspection of the results; hence, the predictions are results of a “modelling” rather than a “fitting” approach.

3.2. Validation of the calculation approach based on ray tracing

For the four configurations described in Section 3.1., a comparison was made between STI measurements and predictions. The result is shown in Fig. 4.

Fig. 4 shows that the ray-tracing approach is able to produce fairly accurate predictions of the STI. A correlation coefficient $r = 0.89$ is satisfactory, especially considering the fact that a measurement error (on average approx. 0.03 STI) was already associated with the STI measurements used in this figure. The average absolute difference between measurement and prediction in Fig. 4 is 0.05 STI; the greatest individual difference for a condition is 0.15 STI. There is also a small systematic difference (bias): on average, the STI is underestimated by 0.019. To put the magnitude of these errors into perspective: the standard deviations of the STI measurements and the STI predictions are both 0.12.

Given that a difference of 0.15 STI is equal to one step on the intelligibility qualification scale given in Table 1, it seems fair to conclude that the adopted prediction approach is sufficiently accurate.

This is further illustrated by Fig. 5, which shows that STI fluctuations as a function of position in the tunnel section (along the length axis) are predicted in relative detail by the calculation procedure. The apparent lesser accuracy in the 75–150 m range is not a systematic effect, but specific to this example; in other configurations, the maximum errors are also observed in different regions.

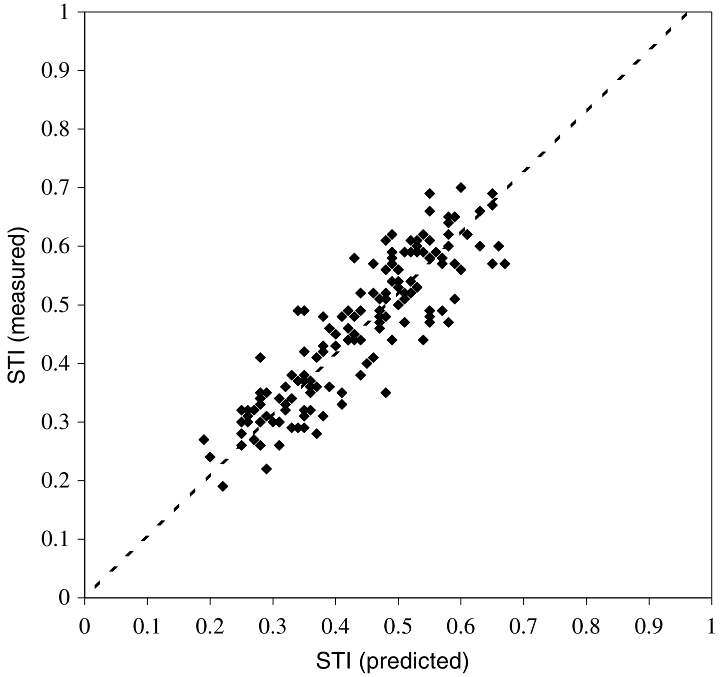

Fig. 4. Correlation between predicted and measured STI values, across a total of 154 measuring points, representing conditions with and without noise, in four tunnels (overall correlation coefficient $r = 0.89$).

3.3. Validation of the calculation approach based on fixed non-linear regression

To show the match between regression-based predictions and measurements, the four available tunnel configurations could again be used in the same way as for the ray-tracing approach. However, the number of independent data points from these four tunnels would be relatively small in relation to the number of fitted parameters. Therefore, ray-tracing simulations were carried out for 13 additional configurations. The measured data of the four original tunnels were also included, bringing the total up to 17. Since the ray-tracing approach was found to be (on average) sufficiently accurate, STI results from this approach are considered an acceptable substitute for actual measurement data. A certain risk is taken in using these ray-tracing predictions, which are not without flaws themselves, as a basis for validation of another prediction technique. However, this drawback is outweighed by the considerable advantage of having a greater data set to work with.

The 17 configurations differed with respect to tunnels dimensions (height and width), number and directivity of loudspeakers and absorption of surfaces in the tunnel. The configurations were chosen to cover the variable ranges given in Table 2 as completely as possible. In all, the model parameters were estimated through regression on 1416 individual values of the effective SNR $S_{\mathrm{ra}}(n)$. This time, only

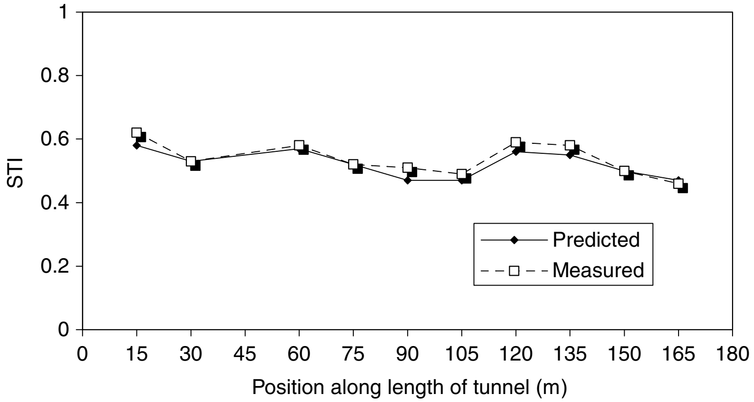

Fig. 5. Measured and predicted STI values for several listener positions (indicated in meters along the length of the tunnel) in tunnel A1 for a condition without noise, at a height of 1.50 m above the road surface.

conditions without noise were considered, since the algorithms for incorporating the effects of noise and non-linear distortions are the same as for the ray-tracing approach (which was found to be sufficiently accurate).

The values for parameters $a(k)$ and $b(k)$ are summarised in Table 3. The value ranges associated with the variables of Eqs. (13) and (14) differ (e.g., 1–7 for octave band n, 125–8000 for linear frequency $f(n)$). In part, this explains why some of the values in Table 3 are numerically much greater then others.

Using Eq. (13) to calculate the $\mathrm{MTF}_{\mathrm{ra}}$ with the parameter values of Table 3, and otherwise using the same algorithms as with the ray-tracing approach, the STI was calculated for 138 representative positions across the 17 different tunnel configurations.

Table 3
Values of parameters $a(k)$ and $b(k)$ derived through fixed non-linear regression on 1416 data points in 17 different tunnel section configurations

k	$a(k)$	$b(k)$
0	-2.497	-1.520
1	-1.698×10^{-2}	-3.652×10^{-2}
2	1.0381	-6.362×10^{-1}
3	8.797×10^{-3}	6.443×10^{-2}
4	1.000	-2.391×10^{-1}
5	-7.6227×10^{-1}	5.788
6	7.526×10^{-3}	-8.481×10^{-2}
7	-4.502×10^{-2}	-7.290×10^{-2}
8	-8.186×10^{-1}	-5.798
9	6.077×10^{-1}	4.550×10^{-7}
10	5.169×10^{-4}	8.753×10^{-4}
11	6.842×10^{-2}	

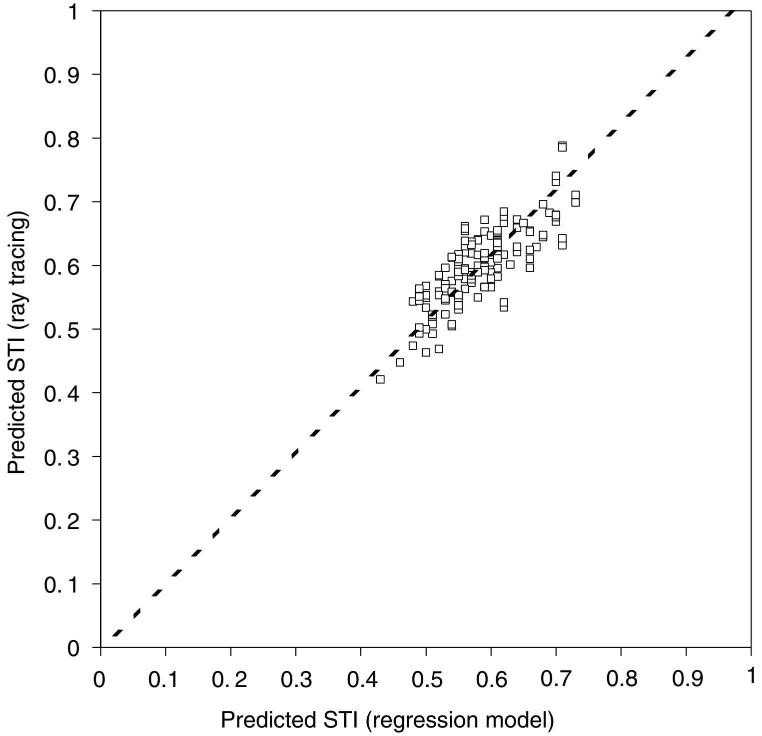

Fig. 6. Correlation between STI values predicted using ray tracing and following the regression approach, across a total of 138 measuring points, all representing realistic conditions without noise, in 17 simulated tunnel environments (overall correlation coefficient $r = 0.81$).

The correlation between these predictions and the predictions obtained through ray tracing are given in Fig. 6.

The mean STI difference between the two prediction types in Fig. 6 is 0.04 STI, the maximum difference is 0.10 STI. For comparison: the standard deviation of both categories of predictions is 0.09. Since no noise conditions were included, no data points are found in the lower end of the STI range. If such noise conditions *are* included then the correlation coefficient is found to increase, but given the fact that the same algorithm is used to incorporate the effects of noise in both types of STI prediction, this is not surprising.

Since the average deviations between both prediction methods are relatively small, and the ray-tracing approach was found to be sufficiently accurate, the same can probably be said for the regression-based approach. If the regression-based predictions are compared to results from the four tunnels (out of 17) for which measurements were available instead of just ray-tracing data, the average error is 0.05 STI, and the correlation coefficient $r = 0.82$. This gives confidence that the data in Fig. 6 do not show a (relatively meaningless) accidental correlation between two prediction procedures that happen to make the same errors.

4. Conclusions and discussion

Both of the proposed STI prediction approaches are suitable for obtaining quantitative insight into PA system performance in tunnels. However, these prediction methods should not be used for comparison of individual STI values against a hard criterion, as sometimes done in STI-measuring practice. If, for instance, each measurement point in a tunnel is required to yield an STI in excess of 0.45, then the average prediction error of 0.05 will probably be considered too high. However, if results are interpreted (or averaged) over multiple conditions and measurement points, then the accuracy of the prediction methods is considered more than adequate.

The strength of the first approach, based on ray tracing, is its general applicability and versatility. Since no parameters were fitted from the data at all, all parameters being specified externally, the high degree of correspondence with (average) measured STI values is very satisfactory.

The statistical regression-based prediction formula was also shown to be sufficiently accurate. Unfortunately, the developed formula and parameters can only be used for this particular traffic tunnel application, and only if the values of all variables fall within the specified ranges. For any new type of application, the best scenario is that the regression parameters need to be re-evaluated. More likely, regression formulae better fitting to the application would have to be devised. The great benefit of the regression-based approach is its computational simplicity. Compared to the ray-tracing approach, the time needed for a typical STI calculation is up to approximately 10^4 times shorter.

Based on the regression approach, a very simple software tool was developed, enabling even non-specialists to quickly evaluate the effects of design changes on the STI. In six steps, the user provides input values for the variables listed in Table 2 (by picking loudspeaker types, surface materials, etc. from a small database). Such a tool targeted to non-specialists in the field of room acoustics and speech intelligibility is useful to evaluate the impact of the usual design choices for tunnel interiors and PA systems. It can be very effective in the early stages of the design process. For more advanced predictions, for instance needed when the tunnel design deviates considerably from the usual shape shown in Fig. 3, it is safer to rely on ray-tracing simulations.

Acknowledgement

This study was in part funded by the Directorate General of Public Works and Water Management, part of the Netherlands Ministry of Transport, Public works and Water Management.

References

[1] Van Wijngaarden SJ, Verhave JA. The influence of fan and traffic noise on speech intelligibility in Dutch traffic tunnels. In: Proceedings of the internoise; 2001.

[2] International Electrotechnical Commission IEC 60268-16 3rd ed. Sound system equipment – Part 16: objective rating of speech intelligibility by Speech Transmission Index. Geneva, Switzerland; 2003.

[3] International Standardization Organization ISO 9921. Ergonomics, Assessment of speech communication. Geneva, Switzerland; 2003.

[4] Steeneken HJM, Houtgast T. A physical method for measuring speech transmission quality. J Acoust Soc Am 1980;67:318–26.

[5] Jacob K, McManus S, Verhave JA, Steeneken HJM. Development of an accurate, handheld, simple-to-use meter for the prediction of speech intelligibility. In: van Wijngaarden SJ, editor. Past present and future of the speech transmission index. Soesterberg: TNO Human Factors; 2002.

[6] Luzzato E, Mouen Makoua D. Application of the W.M.T.F. method for intelligibility prediction in industrial halls. In: Proceedings of the internoise; 1986. p. 905–10.

[7] Van Rietschote HF, Houtgast T, Steeneken HJM. Predicting speech intelligibility in rooms from the modulation transfer function IV: a ray-tracing computer model. Acustica 1981;49:245–52.

[8] Heerema N, Hodgson MR. Empirical models for predicting noise levels, reverberation times and fitting densities in industrial workrooms. Appl Acoust 1999;57(1):51–60.

[9] Hodgson MR, Rempel R, Kennedy S. Measurement and prediction of typical speech and background-noise levels in university classrooms during lectures. J Acoust Soc Am 1999;105(1):226–33.

[10] Hodgson MR. Rating, ranking and understanding acoustical quality in university classrooms. J Acoust Soc Am 2002;112(2):568–75.

[11] Van Wijngaarden SJ, editor. Past, present and future of the speech transmission index. Soesterberg: TNO Human Factors; 2002.

[12] Steeneken HJM. On measuring and predicting speech intelligibility. Doctoral thesis, University of Amsterdam; 1992.

[13] Houtgast T, Steeneken HJM. The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica 1973;28:66–73.

[14] Schroeder MR. Modulation transfer functions: definition and measurement. Acustica 1981;49:179–82.

[15] Houtgast T, Steeneken HJM, Plomp R. Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. Acustica 1980;46:60–72.

[16] Steeneken HJM, Houtgast T. Mutual dependence of the octave-band weights in predicting speech intelligibility. Speech Commun 1999;28:109–23.

[17] Van Rietschote HF, Houtgast T, Steeneken HJM. Predicting speech intelligibility in rooms from the modulation transfer function. IV. A ray-tracing computer model. Acustica 1981;49:245–52.

[18] Christensen CL. Odeon room acoustics program user manual. Department of Acoustic Technology, Technical University of Denmark; 1998.

[19] Statsoft, Inc.. Statistica for windows [computer program manual]. Tulsa: Statsoft; 2000.